

A WEIGHTED DIFFERENCE OF ANISOTROPIC AND ISOTROPIC TOTAL VARIATION FOR RELAXED MUMFORD-SHAH IMAGE SEGMENTATION

Fredrick Park^{*†} Yifei Lou^{*} Jack Xin[†]

^{*†} Whittier College, Mathematics Department, Whittier, CA 90601

^{*}University of Texas Dallas, Department of Mathematical Sciences, Richardson, TX 75080

[†] University of California at Irvine, Mathematics Department, Irvine, CA 92697

ABSTRACT

We propose to incorporate a weighted difference of anisotropic and isotropic total variation (TV) norms into a relaxed formulation of the two phase Mumford-Shah (MS) model for image segmentation. We show results exceeding those obtained by the MS model when using the standard TV norm to regularize partition boundaries. In particular, examples illustrating the qualitative differences between the proposed model and the standard MS one are shown. A fast numerical method is introduced to minimize the proposed model utilizing the difference-of-convex algorithm (DCA) and the primal dual hybrid gradient (PDHG) method.

Index Terms— Total-Variation, Segmentation, Mumford-Shah Functional, Chan-Vese Model, Primal-Dual

1 Introduction

The celebrated Mumford and Shah (MS) model [1] in 2-phase form known as the Chan-Vese (CV) model [2] is one of the most studied and successful models in image processing. The CV model has the following formulation:

$$\min_{\Sigma, c_1, c_2} \text{Per}(\Sigma) + \lambda \int_{\Sigma} (c_1 - f)^2 dx + \int_{\Sigma^c} (c_2 - f)^2 dx. \quad (1)$$

Here, f is a given gray scale image, Σ denotes a region, and Σ^c the outside of that region. The key idea is to minimize the above functional (1) by matching two regions (constants) in the L^2 sense while also minimizing the perimeter of the boundaries between them. The values are designated by c_1 and c_2 and are obtained on the regions Σ and Σ^c respectively. The two regions and values are unknowns and need to be solved for during the minimization. In general, the CV model is difficult to implement in practice. Chan and Vese utilized a level set method formulation and showed successful results. Conventionally, level set methods can be slow to converge due to the need for periodic reinitialization of the level set function to a signed distance one. More recently, Chan et al. [3]

and Bresson et al. [4] proposed convex relaxed formulations of the CV model. The general formulation follows as:

$$\min_{0 \leq u \leq 1} \int_{\Omega} g|\nabla u| dx + \lambda \int_{\Omega} \{(c_1 - f)^2 - (c_2 - f)^2\} u dx \quad (2)$$

where Ω is the image domain and $g(x)$ is an edge detector or set to value 1. The segmented region Σ is realized by taking the upper level set of u in the following way: $\Sigma = \{x : u(x) \geq 1/2\}$. The main idea was to relax the non-convex constraint of u being binary to a convex one. Once the minimization problem (2) is solved, one simply thresholds u to obtain the segmented region. The authors show accurate and successful segmentations along with fast numerical results.

The L_0 norm on the gradient $J(u) = \|\nabla u\|_0$ is construed as the length of partition boundaries in the context of the classical Potts model [5] or two-phase Mumford-Shah segmentation [1]. To circumvent the NP-hard L_0 norm, the convex relaxed approach is to use the L_1 norm on ∇u , see [6]. L_1 norm on the gradient is known as the total-variation (TV) norm [7]. To approximate L_0 , a series of works show that $L_1 - L_2$ is better than L_0 (greedy approaches), L_1 , and L_p in compressive sensing [8]. When applying $L_1 - L_2$ on image gradients, it enforces gradients to be either horizontal or vertical. To account for other gradient directions, a weighted difference, i.e. $L_1 - \alpha L_2$ is considered in the recent work by Lou et al. [9], where α is chosen according to gradient distributions. They show accurate results and provide a convergence proof using a difference-of-convex algorithm (DCA) [10, 11, 12]. $L_1 - 0.5L_2$ is also observed to be numerically more stable compared to $L_1 - L_2$. We also note that the level curves for the convex addition $L_1 + \alpha L_2$ closely resemble those of L_1 while those for $L_1 - \alpha L_2$ are closer to L_0 .

This work is an extension of the work by Lou et al. [9] to the image segmentation problem. To that end, we propose incorporating a weighted difference of anisotropic-isotropic TV into the two-phase relaxed CV model. For the remainder of the paper, we refer to $L_1 - \alpha L_2$ as the weighted difference of anisotropic and isotropic TV:

$$J_{ani} - \alpha J_{iso} = \|u_x\|_1 + \|u_y\|_1 - \alpha \|\sqrt{|u_x|^2 + |u_y|^2}\|_1 \quad (3)$$

Corresponding author email: fepark@whittier.edu.

This work was supported in part by ONR grant N00014-11-1-0602, NSF grant DMS-1222507, NSF grant DMS-1522383, and NSF grant DMS-1522786.

where it is understood that these norms are operating on the gradients of the image. Here, $\alpha \in [0, 1]$ and is chosen based on gradient distributions.

Thus, the key contributions of this paper are:

- Introduce a variant of the relaxed CV model that geometrically better preserves boundaries of regions over the standard CV model.
- Presentation of new numerical schemes utilizing a DCA algorithm with primal dual hybrid gradient (PDHG) methods to solve sub-problems efficiently. DCA has provable convergence properties in certain settings.
- The method can be executed with complete automation.

Related work includes the preceding CV model [2] and convex relaxed variants [3, 4]. A recent hybrid ADMM and dynamic programming to solve the Potts model is introduced by Storath et al. in [13] while convex relaxed approaches are found in Pock et al. [14] and Chambolle et al. [15].

The paper is organized as follows, in section 2 we introduce the proposed model along with the methodology to minimize it. Numerical results are shown in section 3. Lastly, conclusions and future work can be found in section 4.

2 Relaxed Anisotropic-Isotropic Chan-Vese Model (AICV)

We propose a variant of the relaxed CV model proposed by Chan et al. [3] and the CVG model proposed by Bresson et al. [4]. Given an initial image f defined on domain Ω containing a region to be segmented that we denote by Σ , the AICV model seeks to minimize the following:

$$\begin{aligned} & \min_{0 \leq u \leq 1} \int_{\Omega} |u_x(x)| + |u_y(x)| - \alpha |\nabla u(x)| dx \\ & + \lambda \int_{\Omega} \underbrace{\{(c_1 - f(x))^2 - (c_2 - f(x))^2\}}_{r_1(x, c_1, c_2)} u(x) dx \\ & = \min_{0 \leq u \leq 1} F(u, c_1, c_2, \lambda). \end{aligned} \quad (4)$$

The anisotropic-isotropic (AI) term $\int_{\Omega} |u_x| + |u_y| - \alpha |\nabla u|$ denotes a regularization of the region Σ and enforces regularity on the boundary that separates the regions where the two values c_1 and c_2 are taken. We outline the minimization method and subsequent algorithm below.

Assuming that either c_1, c_2 and λ are known or c_1 and c_2 will be solved for during the minimization, we refer to the objective functional $F(u, c_1, c_2, \lambda)$ as $F(u)$ and $r_1(x, c_1, c_2)$ as $r_1(x)$. Then the above AICV model (4) can be written as a minimization of a difference-of-convex (DC) functionals $F(u) = G(u) - H(u)$ in the following manner:

$$\begin{cases} G(u) = \|u_x\|_1 + \|u_y\|_1 + c\|u\|_2^2 + \lambda \langle r_1, u \rangle \\ H(u) = \alpha \|\nabla u\|_{2,1} + c\|u\|_2^2 \end{cases} \quad (5)$$

where c is a small parameter in front of $\|u\|_2^2$ enforcing strong convexity on G and H in the difference. We linearize the

strongly convex term in $H(u)$ and obtain the following iterative scheme:

$$\begin{aligned} u^{n+1} = \arg \min_{0 \leq u \leq 1} & \|u_x\|_1 + \|u_y\|_1 + c\|u\|_2^2 + \lambda \langle r_1, u \rangle \\ & - \alpha \|\nabla u\|_{2,1} - 2c \langle u, u^n \rangle. \end{aligned} \quad (6)$$

The aforementioned minimization associated to the iterative scheme in model (6) has the following equivalent unconstrained split formulation:

$$\begin{aligned} u^{n+1} = \arg \min_{u,v} & \|u_x\|_1 + \|u_y\|_1 + c\|u\|_2^2 + \lambda \langle r_1, v \rangle \\ & + \frac{1}{2\theta} \|u - v\|_2^2 + \beta \langle \nu(v), 1 \rangle - \alpha \|\nabla u\|_{2,1} - 2c \langle u, u^n \rangle. \end{aligned} \quad (7)$$

Here $\nu(\xi) = \max\{0, 2|\xi - \frac{1}{2}| - 1\}$ is an exact penalty given that β is chosen large enough, see [4]. Each DCA subproblem in (7) can be solved by solving the two subproblems:

1. v being fixed, we search for u as a solution of:

$$\begin{aligned} \min_u & \|u_x\|_1 + \|u_y\|_1 - \alpha \|\nabla u\|_{2,1} + \frac{1}{2\theta} \|u - v\|_2^2 \\ & + c\|u\|_2^2 - 2c \langle u, u^n \rangle, \end{aligned} \quad (8)$$

2. u being fixed, we search for v as a solution of:

$$\min_v \frac{1}{2\theta} \|u - v\|_2^2 + \lambda \langle r_1, v \rangle + \beta \int_{\Omega} \nu(v) dx. \quad (9)$$

The solution to (9) is given by a simple shrinkage scheme:

$$v = \min \left\{ \max \left\{ u(x) - \theta \lambda r_1(x, c_1, c_2), 0 \right\}, 1 \right\}. \quad (10)$$

The solution to (8) is obtained by a primal-dual hybrid gradient (PDHG) method by Zhu and Chan [16] which we focus on now. More on the PDHG method and variants thereof can be found in [17].

The minimization problem in (8) reduces to the following min-max problem:

$$\begin{aligned} \min_u \max_{\vec{p} \in \tilde{\mathbf{X}}, \vec{q} \in \mathbf{X}} & \langle u, -\text{div } \vec{p} + \alpha \text{div } \vec{q} \rangle + \frac{1}{2\theta} \|u - v\|_2^2 \\ & + c\|u\|_2^2 - 2c \langle u, u^n \rangle \\ & = \min_u \max_{\vec{p} \in \tilde{\mathbf{X}}, \vec{q} \in \mathbf{X}} \Phi(u, \vec{p}, \vec{q}), \end{aligned} \quad (11)$$

where $\tilde{\mathbf{X}} = \{\vec{p} = \langle p_1, p_2 \rangle : |p_1| \leq 1 \text{ \& } |p_2| \leq 1\}$ and $\mathbf{X} = \{\vec{q} = \langle q_1, q_2 \rangle : |\vec{q}| \leq 1\}$. We will optimize (11) in the following two step manner below.

Step 1. Dual Step

Part A. Fix $u = u^k$ and $\vec{q} = \vec{q}^k$ and apply one step of projected gradient ascent to max problem:

$$\max_{\vec{p} \in \tilde{\mathbf{X}}} \Phi(u^k, \vec{p}, \vec{q}^k). \quad (12)$$

The projected gradient ascent for the maximization (12) is simply:

$$\vec{p}^{k+1} = P_{\tilde{\mathbf{X}}}(\vec{p}^k + \tau_k \nabla u^k) \quad (13)$$

where τ_k is a time step and the projection is obtained by the following operation:

$$P_{\tilde{\mathbf{X}}}(\vec{p}) = \left\langle \frac{p_1}{\max\{|p_1|, 1\}}, \frac{p_2}{\max\{|p_2|, 1\}} \right\rangle. \quad (14)$$

Part B. Fix $u = u^k$ and $\vec{p} = \vec{p}^{k+1}$ and apply one

step of projected gradient ascent to max problem:

$$\max_{\vec{q} \in \mathbf{X}} \Phi(u^k, \vec{p}^{k+1}, \vec{q}). \quad (15)$$

The projected gradient ascent for the maximization (15) is simply:

$$\vec{q}^{k+1} = P_{\mathbf{X}}(\vec{q}^k + \tau_k(-\alpha \nabla u^k)) \quad (16)$$

where τ_k is a time step and the projection is obtained by the following operation:

$$P_{\mathbf{X}}(\vec{q}) = \frac{\vec{q}}{\max\{\|\vec{q}\|, 1\}}. \quad (17)$$

Step 2. Primal Step

Fix $\vec{p} = \vec{p}^{k+1}$ and $\vec{q} = \vec{q}^{k+1}$ and apply one step of gradient descent method to the minimization problem:

$$\min_u \Phi(u, \vec{p}^{k+1}, \vec{q}^{k+1}). \quad (18)$$

The gradient descent associated to the minimization (18) is given by:

$$u^{k+1} = (1 - \theta_k - 2c\theta_k\theta)u^k + \theta_k \left(v - \theta(-\operatorname{div} \vec{p}^{k+1} + \alpha \operatorname{div} \vec{q}^{k+1} - 2cu^n) \right) \quad (19)$$

where θ_k is a time step. The PDHG algorithm for minimizing the unconstrained AICV (7) model is shown in Algorithm 1. We note that when choosing time steps τ_k and θ_k , one can utilize dynamic time stepping to accelerate convergence. Guidelines are given in [17].

Algorithm 1 DCA to minimize unconstrained AICV (7)

Initialization: Pick u^0 , MaxDCA, MaxPDHG, step size τ_k and θ_k , and set $u = 0$.

for 1 **to** MaxDCA **do**

Set $v^0 = \vec{p}^0 = \vec{q}^0 = 0$ and $k \leftarrow 0$

for 1 **to** MaxPDHG **do**

$$v^{k+1} = \min \{ \max \{ u^k - \theta \lambda r_1(x, c_1, c_2), 0 \}, 1 \}$$

$$\vec{p}^{k+1} = P_{\tilde{\mathbf{X}}}(\vec{p}^k + \tau_k \nabla u^k)$$

$$\vec{q}^{k+1} = P_{\mathbf{X}}(\vec{q}^k + \tau_k(-\alpha \nabla u^k))$$

$$u^{k+1} = (1 - \theta_k - 2c\theta_k\theta)u^k + \theta_k(v^{k+1} - \theta(-\operatorname{div} \vec{p}^{k+1} + \alpha \operatorname{div} \vec{q}^{k+1} - 2cu))$$

end for

$$u = u^{k+1}$$

end for

3 Experiments

In this first example we segment a synthetic image to show some properties associated to the proposed model. Observed in Fig.1 (a) is a clean square image while a medium noisy version with initial contour is in (b); Gaussian noise $\sigma = 0.5$. In Fig. 1 (c) the segmentation from the standard CV model is observed where straight boundaries are captured but corners are rounded off. What is well known about the TV norm in regard to denoising also manifests itself in the segmentation

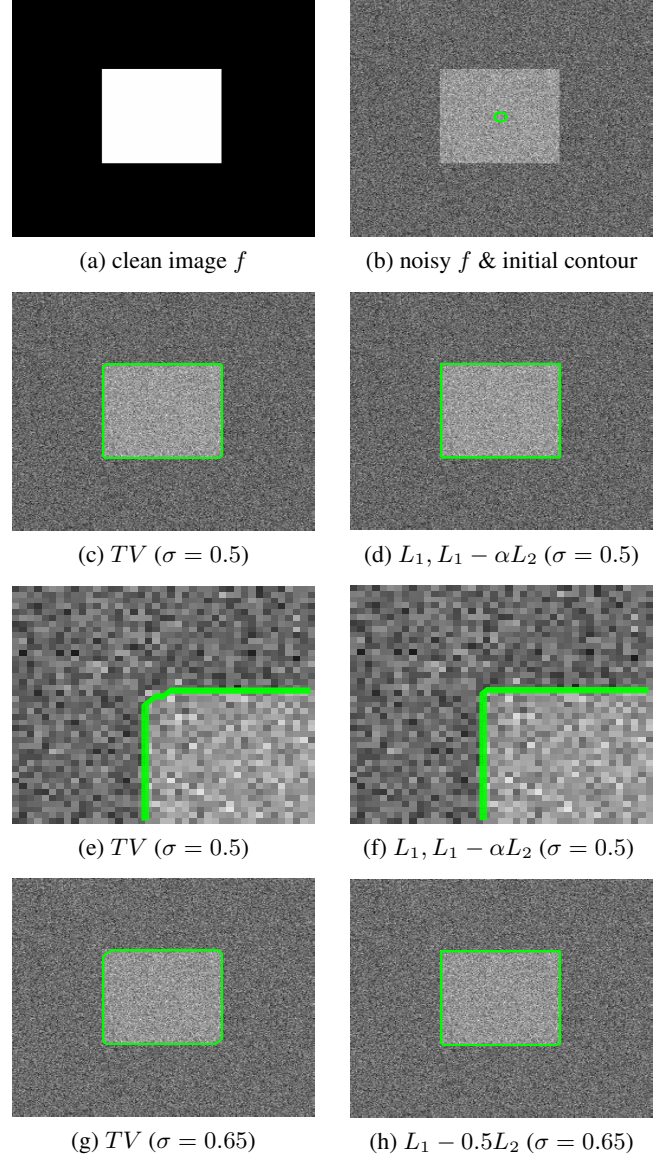


Fig. 1: Square Segmentation Example.

problem, see [18]. The segmentation results for the L_1 and $L_1 - \alpha L_2$ ($\alpha = 1, 0.5$) CV models are seen in Fig. 1 (d) where perfect segmentations are observed. Here, the error $\|f_{\text{clean}} - \{u > 1/2\}\|_2 = 0$ while that for the TV norm model is larger than zero. In Fig 1 (e) and (f) zoom-ins of the top left corner of (c) and (d) respectively where TV in (c) cannot capture the corner. No amount of tuning λ in this case will yield a perfect segmentation. In Fig 1 (g) and (h) we segment the same image under heavier noise, $\sigma = 0.65$. In (g) the TV norm model prominently cuts off the corners while in (h) the proposed model captures them with much higher accuracy.

We now look at a real example and segment a video frame obtained from a sequence under atmospheric turbulence. In Fig. 2 (a) the frame is observed and in (b) a zoomed and processed version from the method in [19] is seen with initial

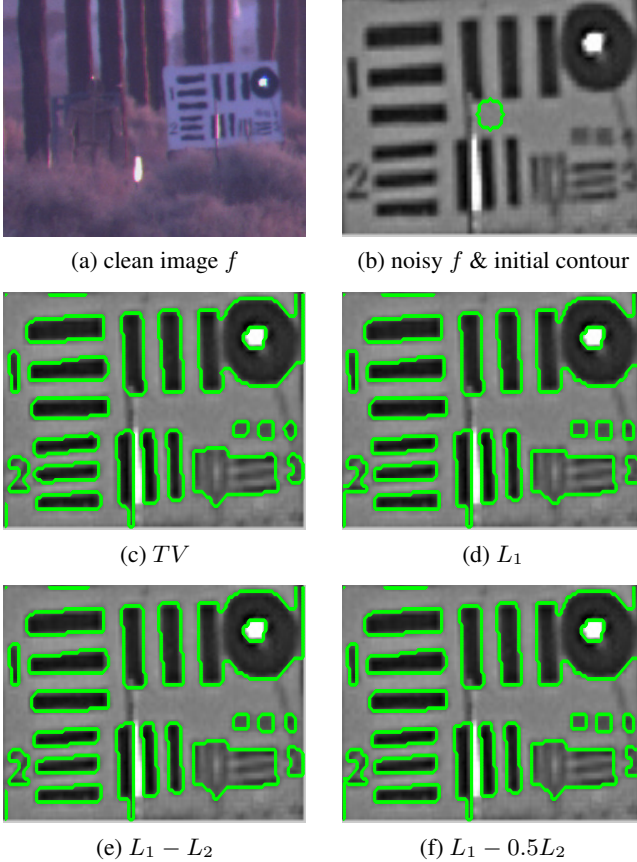


Fig. 2: Sign Segmentation Example.

contour. In (c), (d), (e), and (f) are the segmentation results obtained from the CV model with TV norm, L_1 , $L_1 - L_2$, and $L_1 - 0.5L_2$ norms respectively. All methods capture the large bars, some numbers, and smaller squares. Upon closer inspection, differences are observed like in Fig. 3 (a)–(d) where zoom-ins of the segmentations from the CV model with TV norm, L_1 , $L_1 - L_2$, and $L_1 - 0.5L_2$ norms respectively are seen. For the top 3 squares in (a), the TV norm does not capture the geometry completely. The $L_1 - L_2$ in (c) does a better job than TV while L_1 in (b) and $L_1 - 0.5L_2$ in (d) capture the squares most accurately. The largest contour in (b) from the L_1 norm appears blocky with the lower contour cutting off horizontally versus the smoother version found in (d) from the $L_1 - 0.5L_2$ norm. In Fig. 3 (e) and (f) are zoom-ins of the results from TV and $L_1 - 0.5L_2$ on one large vertical bar where TV cuts off the corner and $L_1 - 0.5L_2$ has straighter geometry keeping the left corner. We deduce that with TV, some geometry is lost but it affords smooth boundaries while L_1 prefers blocky geometry but doesn't favor smooth transitions. $L_1 - L_2$ mimics TV with some blocky geometry preservation. $L_1 - 0.5L_2$ has attributes of L_1 in that it can preserve blocky geometry yet simultaneously keep smoother transitions; a best of both worlds if you will.

In all experiments, λ was tuned to keep as much geometry

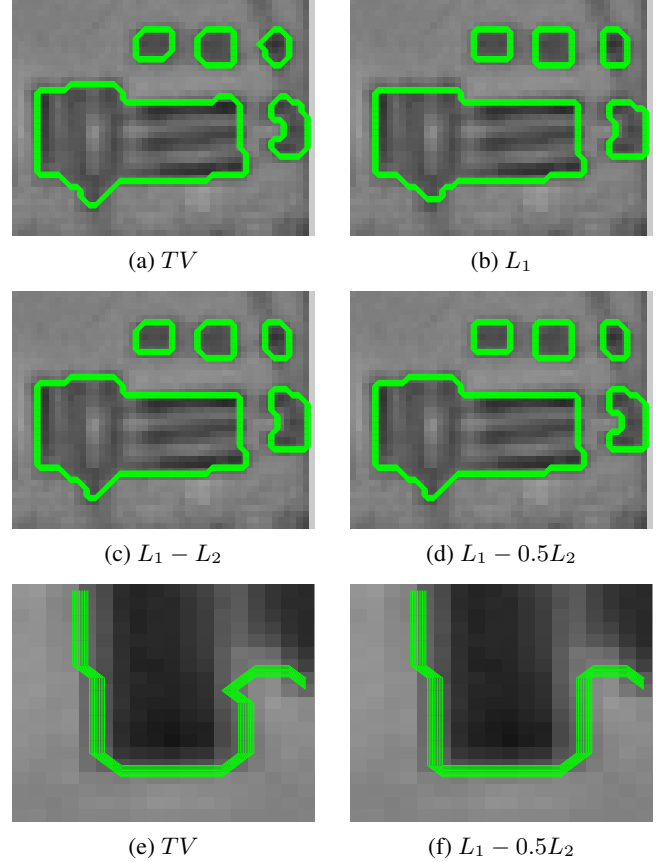


Fig. 3: Sign Segmentation Example Zoom-Ins.

and features in the segmentation. 5000 iterations were used for TV and L_1 , while $\text{MaxDCA} = 5$ and $\text{maxPDHG} = 1000$ for $L_1 - L_2$ and $L_1 - 0.5L_2$ with fixed PDHG time-steps $\tau_k, \theta_k = 1/8$. We fixed the splitting parameter $\theta = 0.1$ in all experiments. Constants c_1 and c_2 were fixed with $c_1 = 1$ and $c_2 = 0$ for the square example while they were $c_1 = 8.5 \times 10^{-2}$ and $c_2 = 7.15 \times 10^{-1}$ for the sign experiment.

4 Conclusions and Future Work

We proposed a relaxed version of the CV model utilizing a difference of convex anisotropic and isotropic TV norms. Moreover, we showed compelling examples illustrating qualitative differences between the proposed model and the standard CV one. A fast algorithm by way of DCA and PDHG was also introduced for its minimization. We note that one of the inner minimizations (8) is non-convex but seems to converge to a minimum consistently in practice. Future work includes a contrast between a weighted difference of convex $L_1 - \alpha L_2$ TV norms and a weighted sum $\alpha L_1 + (1 - \alpha)L_2$ of TV norms i.e. $\alpha \int_{\Omega} (|u_x| + |u_y|) dx + (1 - \alpha) \int_{\Omega} |\nabla u| dx$. Qualitatively, the results are comparable but the level curves for the difference more closely resemble those of the L_0 norm. The advantage of the sum is convexity and this detailed comparison will be found elsewhere. We also plan on proving conver-

gence using the DCA framework for outer iterations on the difference of convex TV norms model. For solving the subproblems, we will be using Bregman [20, 21] inner iterations where the non-convex term is linearized yielding a convex subproblem. A threshold result like in [3] along with extensions of the proposed model to the piecewise smooth MS and multiphase Potts models are also in the works.

5 References

- [1] David Mumford and Shah Jayant, “Boundary detection by minimizing functionals,” in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’85)*, Washington, DC, USA, 1985, CVPR ’85, pp. 137–154, IEEE Computer Society.
- [2] Tony F Chan and Luminita A Vese, “Active contours without edges,” *Image processing, IEEE transactions on*, vol. 10, no. 2, pp. 266–277, 2001.
- [3] Tony F Chan, Selim Esedoglu, and Mila Nikolova, “Algorithms for finding global minimizers of image segmentation and denoising models,” *SIAM journal on applied mathematics*, vol. 66, no. 5, pp. 1632–1648, 2006.
- [4] Xavier Bresson, Selim Esedoglu, Pierre Vanderghenst, Jean-Philippe Thiran, and Stanley Osher, “Fast global minimization of the active contour/snake model,” *Journal of Mathematical Imaging and vision*, vol. 28, no. 2, pp. 151–167, 2007.
- [5] Renfrey Burnard Potts, “Some generalized order-disorder transformations,” in *Mathematical proceedings of the cambridge philosophical society*. Cambridge Univ Press, 1952, vol. 48, pp. 106–109.
- [6] Emmanuel J Candes, Justin K Romberg, and Terence Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [7] Leonid I Rudin, Stanley Osher, and Emad Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.
- [8] Yifei Lou, Penghang Yin, Qi He, and Jack Xin, “Computing sparse representation in a highly coherent dictionary based on difference of L1 and L2,” *Journal of Scientific Computing*, vol. 64, no. 1, pp. 178–196, 2015.
- [9] Yifei Lou, Tiejong Zeng, Stanley Osher, and Jack Xin, “A weighted difference of anisotropic and isotropic total variation model for image processing,” *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1798–1823, 2015.
- [10] Tao Pham-Dinh and Hoai An Le-Thi, “Convex analysis approach to dc programming: Theory, algorithms and applications,” *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.
- [11] Tao Pham-Dinh and Hoai An Le-Thi, “A dc optimization algorithm for solving the trust-region subproblem,” *SIAM Journal on Optimization*, vol. 8, no. 2, pp. 476–505, 1998.
- [12] David J Eyre, “An unconditionally stable one-step scheme for gradient systems,” *Unpublished article*, 1998.
- [13] Martin Storath, Andreas Weinmann, and Laurent Demaret, “Jump-sparse and sparse recovery using potts functionals,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 14, pp. 3654–3666, 2014.
- [14] Thomas Pock, Antonin Chambolle, Daniel Cremers, and Horst Bischof, “A convex relaxation approach for computing minimal partitions,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 810–817.
- [15] Antonin Chambolle, Daniel Cremers, and Thomas Pock, “A convex approach to minimal partitions,” *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, pp. 1113–1158, 2012.
- [16] Mingqiang Zhu and Tony Chan, “An efficient primal-dual hybrid gradient algorithm for total variation image restoration,” *UCLA CAM Report*, pp. 08–34, 2008.
- [17] Antonin Chambolle and Thomas Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [18] Yves Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, vol. 22, American Mathematical Soc., 2001.
- [19] Yifei Lou, Sung Ha Kang, Stefano Soatto, and Andrea L Bertozzi, “Video stabilization of atmospheric turbulence distortion,” *Inverse Problems and Imaging*, vol. 7, no. 3, pp. 839–861, 2013.
- [20] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, “An iterative regularization method for total variation-based image restoration,” *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.
- [21] Tom Goldstein, Xavier Bresson, and Stanley Osher, “Geometric applications of the split bregman method: segmentation and surface reconstruction,” *Journal of Scientific Computing*, vol. 45, no. 1-3, pp. 272–293, 2010.